Tubule Segmentation of Fluorescence Microscopy Images Based on Convolutional Neural Networks With Inhomogeneity Correction

Soonam Lee [a], **Chichen Fu** [a], **Paul Salama** [b], **Kenneth W. Dunn** [c], **and Edward J. Delp** [a]

[a] Video and Image Processing Laboratory (VIPER), School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907, USA
[b] Department of Electrical and Computer Engineering, Indiana University, Indianapolis, Indiana 46202, USA
[c] Division of Nephrology, School of Medicine, Indiana University, Indianapolis, Indiana 46202, USA

Abstract

Fluorescence microscopy has become a widely used tool for studying various biological structures of in vivo tissue or cells. However, quantitative analysis of these biological structures remains a challenge due to their complexity which is exacerbated by distortions caused by lens aberrations and light scattering. Moreover, manual quantification of such image volumes is an intractable and error-prone process, making the need for automated image analysis methods crucial. This paper describes a segmentation method for tubular structures in fluorescence microscopy images using convolutional neural networks with data augmentation and inhomogeneity correction. The segmentation results of the proposed method are visually and numerically compared with other microscopy segmentation methods. Experimental results indicate that the proposed method has better performance with correctly segmenting and identifying multiple tubular structures compared to other methods.

Introduction

Recent advances in fluorescence microscopy imaging, especially two-photon microscopy, have enabled the imaging of cellular and subcellular structures of living tissue [1, 2, 3]. This has resulted in the generation of large datasets of 3D microscopy image volumes, which in turn need automatic image segmentation techniques for quantification. However, the quantitative analyses of these datasets still pose a challenge due to light scattering, distortion created by lens aberrations in different directions, and the complexity of biological structures [4]. The end result is blurry image volumes with poor edge details that become worse in deeper tissue depths.

There have been various techniques developed for segmentation of microscopy images. One widely used class of methods is based on the active contours technique which minimizes an energy functional to fit contours to objects of interest [5, 6]. Early version of active contours [5] generally produced poor segmentation results since the segmentation results are noise sensitive and initial contour dependent. An external energy term which convolves a controllable vector field kernel with an image edge map was presented in [7] to address the noise sensitive problem. Similarly, the Poisson inverse gradient was introduced to determine initial contours locations to segment microscopy images in [8]. Moreover, active contours methodology has been integrated with a region-based segmentation method that poses the segmentation problem as an energy equilibrium problem between foreground and background regions [9]. In addition, this region-based active contours technique was extended to fully utilize 3D information to identify foreground and background voxels [10]. More recently, this 3D region-based active contours was combined with 3D inhomogeneity correction to provide better segmentation since this technique takes into consideration inhomogeneities in volume intensity [11]. Additionally, a new segmentation method known as Squassh that couples image restoration and segmentation using a generalized linear model and Bergman divergence was introduced [12], whereas a method that combined with detecting primitives based on nuclei boundaries and identifying nuclei region using region growing was demonstrated in [13]. Alternatively, combination of multiresolution, multiscale, and region growing methods using random seeds to perform multidimensional segmentation was described in [14].

As indicated above, florescence image segmentation still remains a challenge problem. Tubule, a biological structure with a tubular shape, segmentation is even more challenging since tubular shape and orientation is varied without known patterns. Also, since typical tubular structures have hollow shapes with unclear boundaries, traditional energy minimization based methods such as active contours have failed segmenting tubular structures [15]. There has been some work particularly focusing on tubular structure segmentation. A minimal path based approach was described in [16, 17] where tubule shape is modeled as the envelope of a family of spheres (3D) or disk (2D). Similarly, a new approach for 3D human vessels segmentation and quantification using 3D cylindrical parametric intensity model was demonstrated in [18]. Also, multiple tubule segmentation technique that combined with level set methods and the geodesic distance transform was introduced in [19]. More recently, one method used to segment tubular structures was delineating tubule boundaries followed by ellipse fitting to close the boundaries while considering intensity inhomogeneity [15]. Another method known as Jelly filling [20] utilized adaptive thresholding, component analysis, and 3D consistency to achieve segmentation, whereas a method for tubule boundary segmentation used steerable filters to generate potential seeds from which to grow tubule boundaries followed by tubule/lumen separation and 3D propagation to generate segmented tubules in 3D [21]. Previous methods, however, focused on segmenting bound-

aries of tubule membrane. Since some tubule membranes are not clearly delineated in fluorescence microscopy image volume, finding tubule boundaries may not always result in identifying individual tubule regions.

Convolutional neural network (CNN) has been used to address segmentation problems in biomedical imaging [22]. The fully convolutional network [23] introduced an encoder-decoder architecture for semantic segmentation. U-Net [24] is a 2D CNN based method utilizing this encoder-decoder architecture with connecting intermediate stages of downsampling and upsampling to preserve information. U-Net can be used segment complex biological structure in microscopy images. Similarly, in [25] a U-Net trained on cell objects and contours was used to identify tubular structures. Additionally, a multiple input and multiple output structure based on a CNN for cell segmentation in fluorescence microscopy images was demonstrated in [26]. Also, a nuclei segmentation method that combined with a 2D CNN and a 3D refinement process was introduced in [27].

In this paper, we present a method for segmenting and identifying individual tubular structure based on a combination of intensity inhomogeneity correction, data augmentation, followed by a CNN architecture. Our proposed method is evaluated at object-level metrics as well as pixel-level metrics using manually annotated groundtruth images of real fluorescence microscopy data. Our datasets are comprised of images of a rat kidney labeled with a phalloidin which labels filamentous actin collected using two-photon microscopy. A typical dataset we use in our studies consists of two tissue structures, the base membrane of the tubular structures and the brush border which is generally located interior to proximal tubules. Our goal here is to segment individual tubules enclosed by their membranes.
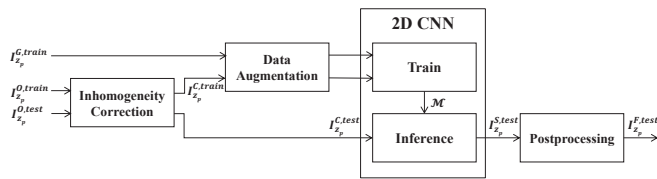
Proposed Method

Figure 1: Block diagram of the proposed segmentation method for tubule segmentation

Figure 1 shows a block diagram of the proposed method. We denote a 3D image volume of size $X \times Y \times Z$ by I, and the pth focal plane image along the z-direction, of size $X \times Y$ pixels, by I_{z_p} where $p \in \{1, \ldots, Z\}$. We also denote the original training and test images in the pth focal plane by $I_{z_p}^{O,train}$ and $I_{z_p}^{O,test}$, respectively. In addition, $I_{z_p}^{G,train}$ and $I_{z_p}^{G,test}$ denote the groundtruth images that are used for training and testing that correspond to $I_{z_p}^{O,train}$ and $I_{z_p}^{O,test}$, respectively. Similarly, $I_{z_p}^{C,train}$ and $I_{z_p}^{C,test}$ denote inhomogeneity corrected training and test images, respectively. Lastly, $I_{z_p}^{S,test}$ denotes the binary segmentation mask generated by our proposed deep learning architecture and $I_{z_p}^{F,test}$ denotes the final segmentation outcome. For example, the 100th original focal plane is denoted as $I_{z_{100}}^{O}$, its corresponding groundtruth image by $I_{z_{100}}^{G}$, the inhomogeneity corrected version by $I_{z_{100}}^{C}$, the binary segmentation mask as $I_{z_{100}}^{S}$, and the final segmentation result

by $I_{z_{100}}^{F}$, respectively.

As shown in Figure 1, our proposed network includes two stages: a training and an inference stage. During the training stage original training images ($I_{z_p}^{O,train}$) have their intensity inhomogeneities corrected ($I_{z_p}^{C,train}$) as a preprocessing step. Since fluorescence microscopy images suffer from intensity inhomogeneity due to non-uniform light attenuation, correcting intensity inhomogeneity helps improve final segmentation results. We then utilize both $I_{z_p}^{C,train}$ and $I_{z_p}^{G,train}$ as inputs to the data augmentation step to increase the number of training image pairs used for training the CNN model, $\mathcal{M}$. During the inference stage inhomogeneity correction is done on the test images ($I_{z_p}^{O,test}$) to obtain $I_{z_p}^{C,test}$. These $I_{z_p}^{C,test}$ are then used to segment tubules with the trained model $\mathcal{M}$.

Intensity Inhomogeneity Correction

Due to non-uniform intensities of fluorescence microscopy where center regions of the focal plane are generally brighter than boundary regions, simple intensity based segmentation methods failed to segment biological structures especially near image boundaries [11]. Our previous work [11] employed a multiplicative model where the original microscopy volume is modeled as

$$I^O = W \circ I^C + N. \tag{1}$$

Here, W and N are a 3D weight array and a zero mean 3D Gaussian noise array, respectively, both of same size as the original microscopy volume. Specifically, W represents weight values for each voxel location that accounts for the degree of intensity inhomogeneity. The $\circ$ operator denotes the Hadamard product representing voxelwise multiplication.

The main idea of the multiplicative model is that an original volume, I^O, is modeled as the product of a 3D inhomogeneity field W with a corrected volume I^C and the product corrupted by additive 3D Gaussian noise N. In [11] an iterative technique to finding W and then correcting for the intensity inhomogeneities based on this model is described.

Our proposed method uses this inhomogeneity correction technique as a preprocessing step for both training and inference. Examples of original and inhomogeneity corrected images are shown in Figure 3a and Figure 3b, respectively.

Data Augmentation

Our training data consists of paired images which are original microscopy images and corresponding manually annotated groundtruth images. Generating manually annotated groundtruth images is a time consuming process and thus impractical when generating large numbers of images. Data augmentation is typically used when the available training data size is relatively small to generate additional groundtruth images [24]. In this paper we utilize an elastic deformation to generate realistic tubular structures with different shapes and orientations. This allows the network to learn various deformed tubular structures, and is particularly useful for analysis of microscopy images especially for tubular structures that appear in varying shapes and orientations [24].

We used elastic deformation by employing a grid of control points located every 64 pixels along the horizontal and vertical directions and displacing these control points randomly within 15
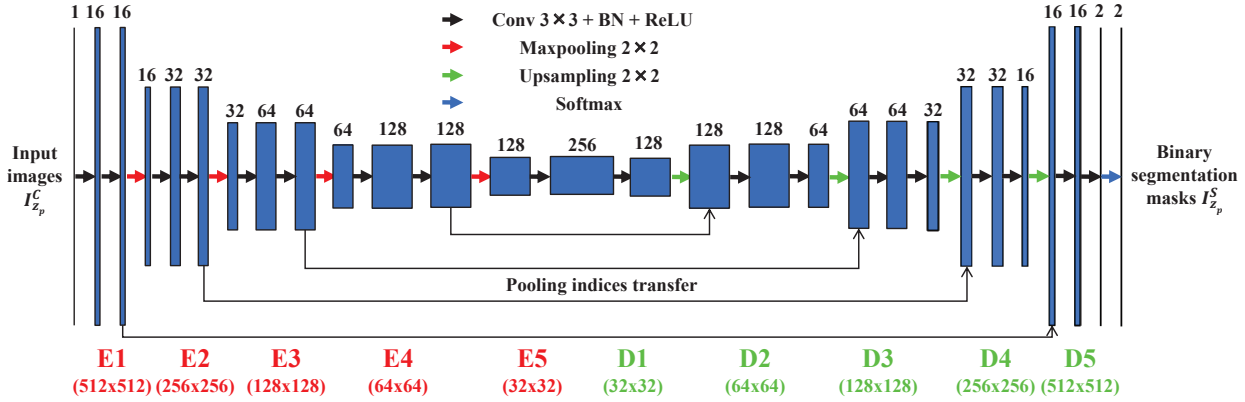
Figure 2: Proposed convolutional neural network architecture

pixels in each direction to generate a deformation field. The deformation field is used to deform the pth focal planes $I_{z_p}^{C,train}$ and $I_{z_p}^{G,train}$ by fitting 2D B-spline basis function to the grid followed by bicubic interpolation [28]. We generated 100 random deformation fields for each image pairs and use them to generate 100 deformed image pairs. Each deformed image is rotated $0°$, $90°$, $180°$, $270°$ to generate four sets of rotated images while preserving the original image size. Each rotated image is then flipped left and right to generate another two sets of images. In our experiment, we manually annotated five pairs of training data during the training stage. Since the elastic deformation uses 100 deformations followed by four rotations and two flips for each deformed image, 4000 pairs of images were generated for training.

Convolutional Neural Network (CNN)

The architecture of our convolutional neural network, shown in Figure 2, consists of 5 encoder layers denoted as $E1$ through $E5$ and 5 decoder layers denoted as $D1$ through $D5$ that are serially connected followed by a softmax layer at the end. Each encoder layer consists of a 3×3 kernel with 1 pixel padding to maintain same image size, a batch normalization step [29] to perform image whitening, followed by a rectifier-linear unit (ReLU). The combination of convolution, batch normalization, and ReLU are performed twice at every encoder. Finally, maxpooling with a stride of 2 is used to reduce dimensionality. This encoder scheme is similar to VGGNet [30] which shrinks the input dimensions but increases the number of filters in the deeper structures. In Figure 2, each encoder's input dimension is indicated in red under the encoder layers. Also note that the number shown above each layer represents the number of filters utilized for training. For example, an input image of size $512 \times 512 \times 1$ is resized to $256 \times 256 \times 16$ at the input to the E2 layer. As the image passes through the all encoder layers, its X and Y dimensions shrink to 32, respectively, but number of filters utilized increases to 256. Therefore, the input to the first decoder layer is of dimension $32 \times 32 \times 256$.

Conversely, each decoder is comprised of two 3×3 kernels with 1 pixel padding, batch normalization, and ReLU. Instead of a maxpooling layer, the decoder has an unmaxpooling layer to upsample the data to increase dimensionality. Note that this upsampling process is a reconstruction process. To achieve better upsampling maxpooling indices from each encoder layer are recorded and transferred to the corresponding same size un-

maxpooling layer ($E1 \rightarrow D5, \ldots, E4 \rightarrow D2$). At the end of the encoder-decoder structure, a softmax classifier layer is utilized to determine whether each pixel location belongs to a tubule or background using a probability map. Note that the output of the softmax layer is of size $512 \times 512 \times 2$ because the final output includes two probability maps corresponding to the two classes: tubule or background. These probability maps are thresholded at 0.5 to produce binary segmentation masks.

During the training stage augmented training images ($I_{z_p}^{C,train}$) are randomly selected and used to train the model $\mathcal{M}$ for each iteration. The segmentation mask is compared with the corresponding groundtruth ($I_{z_p}^{G,train}$) and a loss value is obtained for each iteration. Here, we use a 2D cross entropy loss function that is minimized using stochastic gradient descent (SGD) with a fixed learning rate and a momentum. During the inference stage we use the trained model $\mathcal{M}$ with test images ($I_{z_p}^{C,test}$) to obtain binary segmentation masks ($I_{z_p}^{S,test}$). During the postprocessing step we clean up objects less than γ pixels from $I_{z_p}^{S,test}$ followed by a hole filling operation to obtain final segmentation results ($I_{z_p}^{F,test}$). Note that the hole filling operation assigns a background pixel to a tubule pixel if the background pixel's 4 neighborhood pixels are all tubule pixels.

Experimental Results

The performance of our proposed method was tested on two different datasets:[1] $Dataset - I$ and II. $Dataset - I$ is comprised of $Z = 512$ grayscale images, each of size $X \times Y = 512 \times 512$ pixels, whereas $Dataset - II$ consists of $Z = 821$ grayscale images, each of size $X \times Y = 640 \times 640$ pixels. We selected five different images from $Dataset - I$ and generated corresponding manually annotated groundtruth images to train model $\mathcal{M}$. Our deep learning architecture was implemented in Torch 7 [31] using a fixed learning rate 10^{-5} and a momentum of 0.9. As indicated, 4000 pairs of images were generated by the elastic deformation, rotations, and flips using these five pairs of images. Note that each training data was used as a batch so that 4000 iterations were performed per epoch. We used 200 epochs for training our proposed network. In addition, $\gamma = 100$ was used for the removal of small objects. The performance of the proposed method was evaluated

---

[1] $Dataset - I$ and II were provided by Malgorzata Kamocka of the Indiana Center for Biological Microscopy.

using manually annotated groundtruth images ($I_{z_p}^{G,test}$) at different depths in $Dataset-I$ that were never used during the training stage. For visual evaluation and comparison segmentation results of $I_{z_{100}}$ in $Dataset-I$ using various techniques are presented in Figure 3.
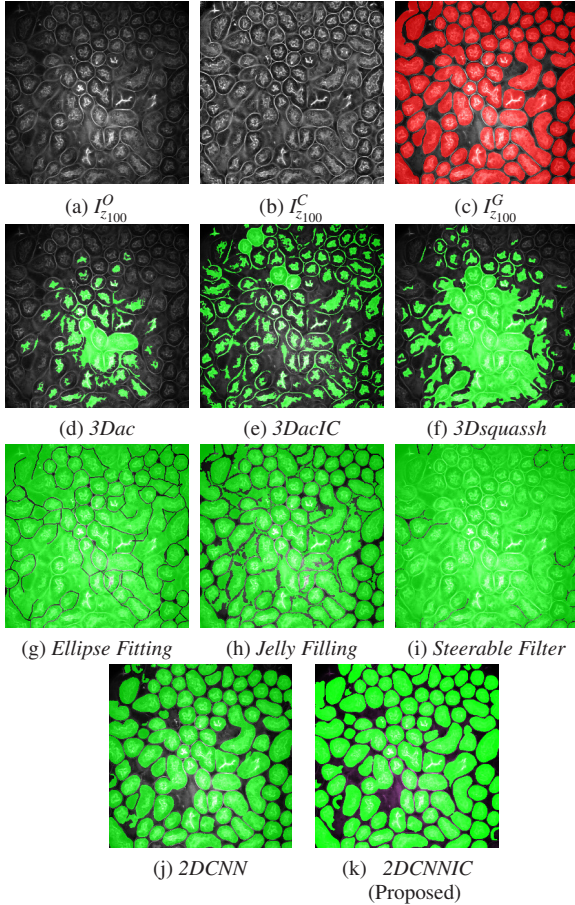

(a) $I_{z_{100}}^{O}$ (b) $I_{z_{100}}^{C}$ (c) $I_{z_{100}}^{G}$

(d) *3Dac*   (e) *3DacIC*   (f) *3Dsquassh*

(g) *Ellipse Fitting*   (h) *Jelly Filling*   (i) *Steerable Filter*

(j) *2DCNN*   (k) *2DCNNIC* (Proposed)

Figure 3: Segmentation results obtained by the proposed method and other methods as well as the corresponding groundtruth data for the 100th image ($I_{z_{100}}$) in $Dataset-I$. Segmentation results are highlighted in green and corresponding groundtruth in red.

Qualitative Evaluation

The first row in Figure 3 displays an original microscopy image ($I_{z_{100}}^{O}$), its inhomogeneity corrected version ($I_{z_{100}}^{C}$), and manually delineated groundtruth ($I_{z_{100}}^{G}$), respectively. For brevity we have omitted the superscript *test* in the notation. The second row shows segmentation results of various 3D methods such as 3D region-based active contours [10] (*3Dac*), 3D active contours with inhomogeneity correction [11] (*3DacIC*), and 3D Squassh presented in [12] (*3Dsquassh*). Similarly, the third row portrays various segmentation methods particularly designed for tubular structure segmentation such as ellipse fitting method presented in [15] (*Ellipse Fitting*), the Jelly filling method in [20] (*Jelly Filling*), and tubule segmentation using steerable filter [21] (*Steerable Filter*). Finally, the last row shows segmentation results of our proposed CNN architecture without inhomogeneity correction [27] (*2DCNN*) and with inhomogeneity correction (*2DCNNIC*).

For visual comparison we highlighted groundtruth regions in red, segmented tubule regions in green, and background in black. As observed in Figure 3, our proposed method appeared to perform better than the other six methods shown in the second and third rows by distinguishing tubules and was similar performance to *2DCNN*. Note that since some methods such as *Ellipse Fitting*, *Jelly Filling*, and *Steerable Filter* only segmented boundaries of tubule structures, tubule interiors were filled in order to perform a fair comparison using connected components with a 4-neighborhood systems. Also, based on the assumption that tubule regions should contain lumen, if a filled region contained lumen pixel, the region was identified as a tubule region. However, if a filled region did not contain any lumen pixels, the region was considered as a background region.

(a) *3Dac*   (b) *3DacIC*   (c) *3Dsquassh*

(d) *Ellipse Fitting*   (e) *Jelly Filling*   (f) *Steerable Filter*
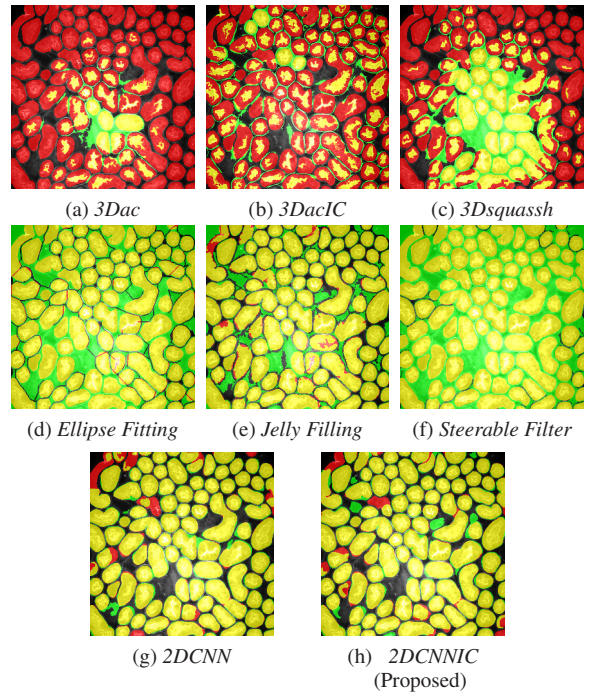
(g) *2DCNN*   (h) *2DCNNIC* (Proposed)

Figure 4: Qualitative evaluation/comparison of tubule segmentation results (shown in green) from the proposed method as well as other methods overlaid onto groundtruth image (shown in red) for $I_{z_{100}}$ belonging to $Dataset-I$

The segmentation results shown in the second row generally missed many tubule regions. More specifically, *3Dac* and *3Dsquassh* could not capture the tubular structures but captured some in the center regions due to the intensity inhomogeneity of microscopy images. *3DacIC* failed to segment tubular structures but captured multiple lumens inside tubules as well as some tubule boundaries. In contrast, the segmentation results displayed in the third row showed falsely detected tubules. The main reason is that these tubule segmentation methods focused only on detecting boundaries of tubular structures. In particular, due to weak/blurry edges of fluorescence microscopy images, many boundaries were not continuous causing the filling operation to overflow from one tubule to another or to the background regions. The segmentation results using the CNN generally successfully segmented and identified each tubule region.

Figure 4 provides an alternative way to show the segmen-

Method	$I_{z_{100}}$ of the *Dataset − I*						$I_{z_{200}}$ of the *Dataset − I*					
	PA	Type-I	Type-II	F1	OD	OH	PA	Type-I	Type-II	F1	OD	OH
3Dac [10]	37.74%	**3.31%**	58.95%	0.90%	20.86%	95.72	38.98%	**2.72%**	58.30%	1.83%	20.74%	125.69
3DacIC [11]	42.92%	8.06%	49.02%	0.86%	36.45%	35.07	44.58%	4.84%	50.59%	0.00%	38.02%	30.41
3Dsquassh [12]	47.02%	11.80%	41.18%	1.83%	11.64%	223.34	48.37%	9.55%	42.09%	1.94%	14.34%	181.58
Ellipse Fitting [15]	76.17%	22.79%	1.04%	61.15%	47.10%	144.28	76.11%	22.98%	0.91%	48.48%	29.34%	303.34
Jelly Filling [20]	83.91%	13.36%	2.73%	81.82%	71.58%	52.93	81.76%	15.38%	2.86%	74.53%	60.73%	76.26
Steerable Filter [21]	70.98%	28.98%	**0.04%**	9.90%	5.32%	455.83	71.00%	28.97%	**0.03%**	4.12%	4.00%	521.83
2DCNN [27]	**90.57%**	5.25%	4.17%	91.49%	90.09%	13.28	86.92%	5.64%	7.44%	86.96%	87.10%	16.80
2DCNNIC (Proposed)	90.04%	6.44%	3.52%	**92.63%**	**90.12%**	**11.95**	**88.66%**	5.77%	5.57%	**90.61%**	**89.65%**	**11.76**

Table 1: Quantitative evaluation of the proposed method and other known methods in terms of Pixel Accuracy (PA), Type-I error, Type-II error, F1 score, Dice Index (OD), and Hausdorff Distance (OH)

tation results. In particular, yellow regions correspond to true positives which are pixel locations that are identified as tubules in both the groundtruth and segmentation results. Green regions correspond to false positives which are pixel locations that are identified as background in groundtruth but tubules in segmentation results. Similarly, red pixels correspond to false negatives, namely pixel locations identified as tubules in the groundtruth but background in segmentation results, and black pixel regions correspond to true negative that are identified as background in both groundtruth and segmentation results. The green regions indicate Type-I error (false alarm) regions and the red regions represent Type-II error (miss) regions. As observed from Figure 4, the segmentation results in the first row contained large red regions which mean large regions of tubules were missed. Conversely, the segmentation results shown in the second row contained many green regions indicating many background regions were falsely segmented as tubule regions. In contrast, the segmentation results in the third row had reasonably small green regions and red regions which indicate that the deep learning based segmentation results had higher pixel accuracy with relatively low Type-I and Type-II errors.

Quantitative Evaluation

In addition to the qualitative evaluation, quantitative metrics for evaluating the proposed method's segmentation accuracy of objects were utilized. In particular, we used pixel-based and object-based metrics. In the pixel-based metric, the pixel accuracy (PA), Type-I error, and Type-II error of pixel segmentation were obtained based on the manually annotated groundtruth images. Here, PA, Type-I, and Type-II are defined as below:

$$PA = \frac{N_{tp}^p + N_{tn}^p}{N_{total}^p}, \quad Type-I = \frac{N_{fp}^p}{N_{total}^p}, \quad Type-II = \frac{N_{fn}^p}{N_{total}^p} \quad (2)$$

where N_{tp}^p, N_{tn}^p, N_{fp}^p, and N_{fn}^p are defined to be the number of segmented pixels that were labeled as true positives, true negatives, false positives, false negatives, respectively. N_{total}^p denotes the total number of pixels in a image. These three pixel-based metrics obtained for 8 different segmentation results are provided in Table 1. As shown in Figure 4, Type-II errors of the first three methods (*3Dac*, *3DacIC*, *3Dsquassh*) were much higher compared to other methods. Similarly, Type-I errors of next three methods (*Ellipse Fitting*, *Jelly Filling*, *Steerable Filter*) were much higher than those of the other methods. However, *2DCNN* and *2DCNNIC* had high PA and relatively low Type-I and Type-II errors.

In addition, our segmentation methods were evaluated using object-based criteria described in the *2015 MICCAI Grand Segmentation Challenge* [25, 32] namely: the F1 score metric, the Dice Index, and the Hausdorff Distance.

The F1 score metric is a measure of the segmentation/detection accuracy of individual objects. The evaluation of the F1 score metric is based on two metrics, precision P and recall R. Denoting the number of tubules correctly identified by N_{tp}^o, the number of objects that are non-tubules but identified as tubules by N_{fp}^o, and the number of tubules that are not correctly identified as tubules by N_{fn}^o, respectively, then precision P and recall R are obtained as [25]

$$P = \frac{N_{tp}^o}{N_{tp}^o + N_{fp}^o} \quad \text{and} \quad R = \frac{N_{tp}^o}{N_{tp}^o + N_{fn}^o}. \quad (3)$$

Given the values of P and recall R, the F1 is found using

$$F1 = \frac{2PR}{P+R}. \quad (4)$$

It is to be noted that a tubule segmented by the proposed method (or any other method for that matter) that overlaps at least 50% with its corresponding manually annotated tubule is labeled as a true positive and added to the count of the true positives (N_{tp}^o), otherwise it is considered as a false positive and added to the count of the false positives (N_{fp}^o). Similarly, a manually annotated tubule that has no corresponding segmented tubule or overlaps less than 50% with segmented tubular regions is considered to be a false negative and added to the count of the false negatives (N_{fn}^o).

As mentioned above a second metric used to evaluate segmentation accuracy is the Dice Index (OD). The Dice Index [33] is a measure of similarity between two sets of samples. In our case, the two sets of samples are the sets of voxels belonging to a manually annotated tubule denoted by G, and the set of voxels belonging to a segmented tubule denoted by S. The Dice Index between G and S is defined as

$$D(G,S) = \frac{2|G \cap S|}{|G| + |S|} \quad (5)$$

where $|\cdot|$ denotes set cardinality which in this case will be the number of voxels belonging to an object. A higher value of the Dice Index indicates better segmentation match/results relative to the groundtruth data. A practical way of evaluating the Dice Index

for segmented objects is described in [32] and is given by

$$D(G,S) = \frac{1}{2}\left[\sum_{i=1}^{n_S} w_i D(G_i,S_i) + \sum_{j=1}^{n_G} \tilde{w}_j D(\tilde{G}_j,\tilde{S}_j)\right] \qquad (6)$$

where

$$w_i = |S_i|/|\sum_{p=1}^{n_S}|S_p|, \quad \tilde{w}_j = |\tilde{G}_j|/|\sum_{q=1}^{n_G}|\tilde{G}_q|. \qquad (7)$$

In Eq (6), S_i denotes the ith tubule ($i \in \{1,\ldots,n_S\}$) obtained by a segmentation method and G_i denotes a manually annotated tubule that is maximally matched with S_i. Similarly, $\tilde{G}_j$ denotes the jth tubule ($j \in \{1,\ldots,n_G\}$) identified in the groundtruth data and $\tilde{S}_j$ denotes a segmented tubule that is maximally matched with $\tilde{G}_j$. Finally, n_S and n_G denote the total number of segmented and manually annotated tubules, respectively. The first summation term in Eq (6) represents how well each groundtruth tubule overlaps with its segmented counterpart, whereas the second summation term represents how well each segmented tubule overlaps with its manually annotated counterpart. The terms w_i and $\tilde{w}_j$ which are used to weight the summation terms represent the fraction of the space that each tubule S_i and $\tilde{G}_j$ occupies within the entire tubule region, respectively.

While the Dice Index measures segmentation accuracy, a third metric, the Hausdorff Distance (OH), is needed to evaluate shape similarity. The Hausdorff Distance [34], $H(G,S)$, between a segmented tubule S and its manually annotated counterpart G, is defined to be

$$H(G,S) = \max\{\sup_{x \in G}\inf_{y \in S}||x-y||_2, \ \sup_{y \in S}\inf_{x \in G}||x-y||_2\}. \qquad (8)$$

Here, $||x-y||_2$ denotes the Euclidean distance between a pair of pixels x and y. Based on Eq (8), the Hausdorff Distance obtains the maximum distance among all pairs of voxels on the boundaries of S and G. Therefore, a smaller value of the Hausdorff Distance indicates a higher similarity in shape between the boundaries of S and G. As done above (see Eq (6)), a practical way of finding the Hausdorff Distance between a segmented tubule S and its manually annotated counterpart G is given by [32]:

$$H(G,S) = \frac{1}{2}\left[\sum_{i=1}^{n_S} w_i H(G_i,S_i) + \sum_{j=1}^{n_G} \tilde{w}_j H(\tilde{G}_j,\tilde{S}_j)\right] \qquad (9)$$

where the parameters w_i and $\tilde{w}_j$ are defined in Eq (7).

The performance of the proposed method and other methods based on the F1 score, OD, and OH metrics were obtained and tabulated in Table 1. As mentioned above higher values of F1 and OD are considered to be indicators of better segmentation results. In contrast, lower values of OH indicate better segmentation result. As can be seen in Table 1, our proposed method outperformed all the other segmentation methods against which the proposed method is being evaluated. In particular, *3Dac*, *3DacIC*, *3Dsquassh*, and *Steerable Filter* had low F1 scores since these segmentation methods had large Type-I or Type-II errors. Similarly, all of the methods except for *2DCNN* suffered from low OD and high OH values. In particular, since the segmentation results of *3Dsquassh*, *Ellipse Fitting*, and *Steerable Filter* failed to

(a) $I^O_{z_{100}}$ of *Dataset − I* (b) $I^O_{z_{150}}$ of *Dataset − I* (c) $I^O_{z_{200}}$ of *Dataset − I*

(d) $I^F_{z_{100}}$ of *Dataset − I* (e) $I^F_{z_{150}}$ of *Dataset − I* (f) $I^F_{z_{200}}$ of *Dataset − I*

(g) $I^O_{z_{50}}$ of *Dataset − II* (h) $I^O_{z_{150}}$ of *Dataset − II* (i) $I^O_{z_{250}}$ of *Dataset − II*

(j) $I^F_{z_{50}}$ of *Dataset − II* (k) $I^F_{z_{150}}$ of *Dataset − II* (l) $I^F_{z_{250}}$ of *Dataset − II*
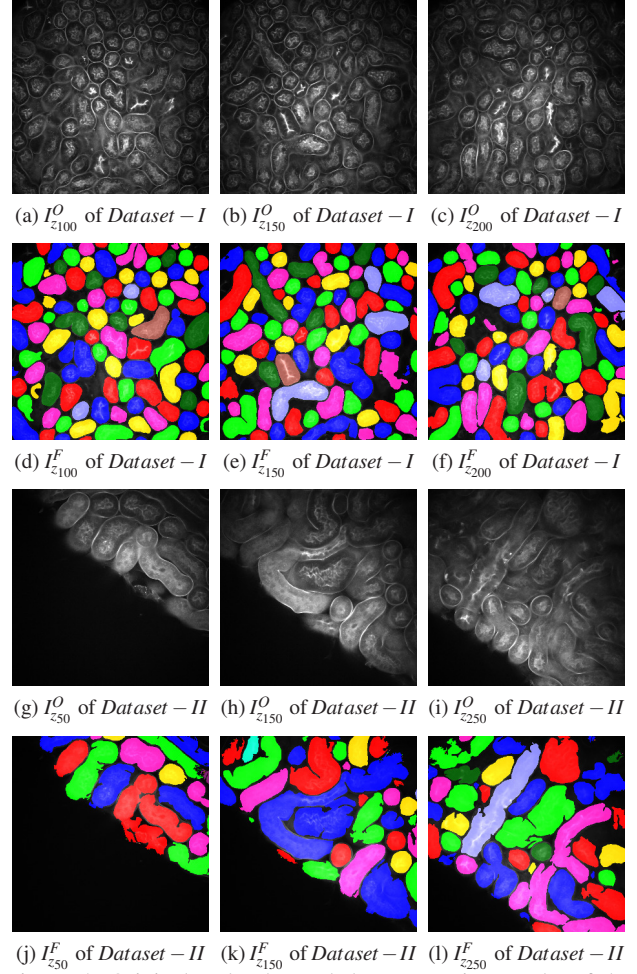
Figure 5: Original and color coded segmentation results of the proposed method on different depth of *Dataset − I* and *Dataset − II* using same trained model $\mathcal{M}$

distinguish most of the individual tubules, they exhibited low OD and high OH values. Note that *3DacIC* had relatively low OH *and* low OD values since it segmented some tubule boundaries as well as some partial regions (lumen) inside the tubules. Lastly, the use of intensity inhomogeneity correction in the proposed method improved its performance relative to that of *2DCNN*.

For visual evaluation we provide the segmentation results of the proposed method using two different datasets: *Dataset − I* and *Dataset − II*, sampled at different depths within the volumes. The first row shows original microscopy images $I^O_{z_{100}}$, $I^O_{z_{150}}$, and $I^O_{z_{200}}$ from *Dataset − I* and the second row displays the segmentation results corresponding to the first row. To better visualize the segmentation results, we highlighted individual tubules with different colors and overlaid them onto the original microscopy images. Similarly, the third row exhibits original microscopy images $I^O_{z_{50}}$, $I^O_{z_{150}}$, and $I^O_{z_{250}}$ from *Dataset − II*. Their corresponding segmentation results are shown in the fourth row. Note that the model $\mathcal{M}$ which was trained on *Dataset − I* was used for *Dataset − II* during the inference stage. Although the shape, size, and orientation of tubular structures presented in *Dataset − II* are all different from *Dataset − I*, the proposed method can still successfully segment

and identify individual tubules presented in *Dataset − II* as well as individual tubules in *Dataset − I*.

Conclusions

This paper presented a tubular structure segmentation method that uses inhomogeneity correction, data augmentation, a convolutional neural network, and postprocessing. The qualitative and quantitative results indicate that the proposed method can successfully segment and identify individual tubules compared to other segmentation methods. In the future, we plan to utilize 3D information generated from realistic 3D synthetic tubular structures to improve segmentation results as well as reduce manual annotation work.

Acknowledgment

References

[1] R. K. P. Benninger, M. Hao, and D. W. Piston, "Multi-photon excitation imaging of dynamic processes in living cells and tissues," *Reviews of Physiology Biochemistry and Pharmacology*, vol. 160, pp. 71–92, April 2008.

[2] F. Helmchen and W. Denk, "Deep tissue two-photon microscopy," *Nature Methods*, vol. 2, no. 12, pp. 932–940, December 2005.

[3] K. W. Dunn, R. M. Sandoval, K. J. Kelly, P. C. Dagher, G. A. Tanner, S. J. Atkinson, R. L. Bacallao, and B. A. Molitoris, "Functional studies of the kidney of living animals using multicolor two-photon microscopy," *American Journal of Physiology-Cell Physiology*, vol. 283, no. 3, pp. C905–C916, September 2002.

[4] D. B. Murphy and M. W. Davidson, *Fundamentals of Light Microscopy and Electronic Imaging*, Wiley-Blackwell, Hoboken, NJ, 2nd edition, 2012.

[5] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, January 1988.

[6] R. Delgado-Gonzalo, V. Uhlmann, D. Schmitter, and M. Unser, "Snakes on a plane: A perfect snap for bioimage analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 1, pp. 41–48, January 2015.

[7] B. Li and S. T. Acton, "Active contour external force using vector field convolution for image segmentation," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2096–2106, August 2007.

[8] B. Li and S. T. Acton, "Automatic active model initialization via Poisson inverse gradient," *IEEE Transactions on Image Processing*, vol. 17, no. 8, pp. 1406–1420, August 2008.

[9] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, February 2001.

[10] K. S. Lorenz, P. Salama, K. W. Dunn, and E. J. Delp, "Three dimensional segmentation of fluorescence microscopy images using active surfaces," *Proceedings of the IEEE International Conference on Image Processing*, pp. 1153–1157, September 2013, Melbourne, Australia.

[11] S. Lee, P. Salama, K. W. Dunn, and E. J. Delp, "Segmentation of fluorescence microscopy images using three dimensional active contours with inhomogeneity correction," *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 709–713, April 2017, Melbourne, Australia.

[12] G. Paul, J. Cardinale, and I. F. Sbalzarini, "Coupling image restoration and segmentation: A generalized linear model/Bregman perspective," *International Journal of Computer Vision*, vol. 104, no. 1, pp. 69–93, March 2013.

[13] S. Arslan, T. Ersahin, R. Cetin-Atalay, and C. Gunduz-Demir, "Attributed relational graphs for cell nucleus segmentation in fluorescence microscopy images," *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 1121–1131, June 2013.

[14] G. Srinivasa, M. C. Fickus, Y. Guo, A. D. Linstedt, and J. Kovacevic, "Active mask segmentation of fluorescence microscope images," *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1817–1829, August 2009.

[15] S. Lee, P. Salama, K. W. Dunn, and E. J. Delp, "Boundary fitting based segmentation of fluorescence microscopy images," *Proceedings of the IS&T/SPIE Conference on Imaging and Multimedia Analytics in a Web and Mobile World 2015*, pp. 940805–1–10, February 2015, San Francisco, CA.

[16] F. Benmansour and L. D. Cohen, "Tubular structure segmentation based on minimal path method and anisotropic enhancement," *International Journal of Computer Vision*, vol. 92, no. 2, pp. 192–210, March 2010.

[17] H. Li and A. Yezzi, "Vessels as 4-D curves: Global minimal 4-D paths to extract 3-D tubular surfaces and centerlines," *IEEE Transactions on Medical Imaging*, vol. 26, no. 9, pp. 1213–1223, September 2007.

[18] S. Worz and K. Rohr, "A new 3D parametric intensity model for accurate segmentation and quantification of human vessels," *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 491–499, September 2004, Saint-Malo, France.

[19] A. Fakhrzadeh, E. Sporndly-Nees, L. Holm, and C. L. Luengo Hendriks, "Analyzing tubular tissue in histopathological thin sections," *Proceedings of the IEEE International Conference on Digital Image Computing Techniques and Applications*, pp. 1–6, December 2012, Fremantle, WA.

[20] N. Gadgil, P. Salama, K. W. Dunn, and E. J. Delp, "Jelly filling segmentation of fluorescence microscopy images containing incomplete labeling," *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 531–535, April 2016, Prague, Czech Republic.

[21] D. J. Ho, P. Salama, K. W. Dunn, and E. J. Delp, "Boundary segmentation for fluorescence microscopy using steerable filters," *Proceedings of the SPIE Conference on Medical Imaging*, pp. 10133–1–11, February 2017, Orlando, FL.

[22] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sanchez, "A survey on deep learning in medical image analysis," *arXiv preprint arXiv:1702.05747*, pp. 1–34, February 2017.

[23] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, April 2017.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*, vol. 9351, pp. 234–241, October 2015, Munich, Germany.

[25] H. Chen, X. Qi, L. Yu, and P. A. Heng, "DCAN: Deep contour-aware networks for accurate gland segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2487–2496, June 2016, Las Vegas, NV.

[26] S. E. A. Raza, L. Cheung, D. Epstein, S. Pelengaris, M. Khan, and N. M. Rajpoot, "MIMO-NET: A multi-input multi-output convolutional neural network for cell segmentation in fluorescence microscopy images," *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 337–340, April 2017, Melbourne, Australia.

[27] C. Fu, D. J. Ho, S. Han, P. Salama, K. W. Dunn, and E. J. Delp, "Nuclei segmentation of fluorescence microscopy images using convolutional neural networks," *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 704–708, April 2017, Melbourne, Australia.

[28] K. S. Lorenz, P. Salama, K. W. Dunn, and E. J. Delp, "Digital correction of motion artefacts in microscopy image sequences collected from living animals using rigid and nonrigid registration," *Journal of Microscopy*, vol. 245, no. 2, pp. 148–160, February 2012.

[29] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *Proceedings of the International Conference on Machine Learning*, vol. 37, pp. 448–456, July 2015, Lille, France.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, pp. 1–14, April 2015.

[31] R. Collobert, K. Kavukcuoglu, and C. Farabet, "Torch7: A matlab-like environment for machine learning," *Proceedings of the BigLearn workshop at the Neural Information Processing Systems*, pp. 1–6, December 2011, Granada, Spain.

[32] K. Sirinukunwattana, J. P. W. Pluim, H. Chen, X. Qi, P. A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, A. Bohm, O. Ronneberger, B. B. Cheikh, D. Racoceanu, P. Kainz, M. Pfeiffer, M. Urschler, D. R. J. Snead, and N. M. Rajpoot, "Gland segmentation in colon histology images: The glas challenge contest," *Medical Image Analysis*, vol. 35, no. 1, pp. 489–502, January 2017.

[33] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, July 1945.

[34] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, September 1993.

Author Biography

Soonam Lee received the B.S. (Summa Cum Laude) degree in Electrical and Computer Engineering from Hanyang University, Seoul, Korea in 2008. He received the M.S. degree in Electrical Engineering and Computer Science from University of Michigan, MI in 2012. He also received the M.S. degree in Mathematics from University of Michigan, MI in 2012. He is currently pursuing Ph.D. at the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN. His research interests include image processing, computer vision, machine learning, deep learning and biomedical imaging area.

Chichen Fu received his B.S. in Electrical Engineering from Purdue University in 2014. He is currently pursuing his Ph.D. in Electrical and Computer Engineering from Purdue University. His research interests lie in machine learning, image processing and computer vision.

Paul Salama is Professor of Electrical and Computer Engineering and Assistant Dean of Graduate Programs at the Purdue School of Engineering and Technology at the Indiana University Purdue University, Indianapolis. His research interests include image and video processing, image analysis, image and video compression, medical imaging, and machine learning. Dr. Salama is a Senior Member of the IEEE and member of the SPIE.

Kenneth W. Dunn was born in Redwood City, California. He received his Ph.D. at the State University of New York at Stony Brook, and completed his post-doctoral training at Columbia University, where he was subsequently appointed as an Assistant Professor. In 1995 he took a position in the Nephrology Division in the School of Medicine at Indiana University, where he subsequently rose to the rank of Professor of Medicine and Biochemistry. His research interests include epithelial cell biology, drug-induced liver injury and quantitative light microscopy. In his role as Director of the Indiana Center for Biological Microscopy, he has overseen the development quantitative and intravital microscopy at the IU School of Medicine, directing core services for NIH program projects in Hematology, Diabetes and Nephrology.

Edward J. Delp was born in Cincinnati, Ohio. He is currently The Charles William Harrison Distinguished Professor of Electrical and Computer Engineering and Professor of Biomedical Engineering at Purdue University. His research interests include image and video processing, image analysis, computer vision, image and video compression, multimedia security, medical imaging, multimedia systems, communication and information theory. Dr. Delp is a Life Fellow of the IEEE, a Fellow of the SPIE, a Fellow of IS&T, and a Fellow of the American Institute of Medical and Biological Engineering.